

**КЛАСТЕРИЗАЦИЯ РОССИЙСКИХ ВУЗОВ В ФАКТОРНОМ ПРОСТРАНСТВЕ
ВСТУПИТЕЛЬНЫХ ИСПЫТАНИЙ ПО НАПРАВЛЕНИЮ ПОДГОТОВКИ МАТЕМАТИКА**

Д.А. Новосельцева

Научный руководитель: доцент, к ф.-м. н. А.А. Михальчук

Национальный исследовательский Томский политехнический университет,

Россия, г.Томск, пр. Ленина, 30, 634050

E-mail: dary_2503@mail.ru**CLUSTERING RUSSIAN UNIVERSITIES IN THE FACTOR SPACE OF ENTRANCE
EXAMINATIONS BY FIELD OF EDUCATION MATHEMATICS**

D.A. Novoseltseva

Scientific Supervisor: As. Prof., PhD A.A. Mikhalechuk

Tomsk Polytechnic University, Russia, Tomsk, Lenin str., 30, 634050

E-mail: dary_2503@mail.ru

Abstract. This article was performed multivariate statistical analysis of experimental data quality entrants in Russian universities by field of education mathematics (FEM). Constructed 4-factor model for the 9 indicators of quality set in the university. Constructed 12 cluster model of 119 Russian universities FEM, which allows to allocate groups of universities, homogeneous sets of factorial indicators. The results of the statistical analysis may be taken into account in decision-making in the ongoing reform of higher education.

В данной работе, выполнен многомерный статистический анализ экспериментальных данных качества набора абитуриентов в ВУЗы по направлению подготовки математика (НПМ), включающему в себя, в частности, подготовку по прикладной математике и информатике [1]. В качестве базы данных взяты данные 119 ВУЗов по следующим 9-ти показателям: $ЕГЭ_{15}$ - средний балл ЕГЭ зачисленных в 2015 году, $ЕГЭ_{14}$ - средний балл ЕГЭ зачисленных в 2014 году, $ЕГЭ_{ид}$ - средний балл ЕГЭ зачисленных в 2015 году с учетом индивидуальных достижений, N - количество студентов, зачисленных на бюджетные места, N_c - количество студентов, зачисленных на бюджетные места по результатам конкурса, N_{TS} - количество студентов, зачисленных на целевые места, N_L - количество студентов, зачисленных на бюджетные места по льготам, N_o - количество студентов, зачисленных на бюджетные места по результатам олимпиады, $No\%$ - долевое количество студентов (в % от общего количества бюджетных мест), принятых по олимпиадам. Из анализа были исключены те ВУЗы, в которых численность поступивших на бюджетные места по направлению подготовки Математика не превышала 10 человек. Таким образом, созданная в MSExcel база данных впоследствии использовалась в пакете Statistica[2] для статистического анализа данных.

В рамках корреляционного анализа обнаружены высоко значимые (на уровне значимости $p < 0,0005$) положительная корреляционная зависимость между показателями в группах $\{ЕГЭ_{15}, ЕГЭ_{14} \text{ и } ЕГЭ_{ид}\}, \{N, N_c \text{ и } N_L\}$, а также $\{No \text{ и } No\%\}$ (рис.1).

С помощью факторного анализа методом главных компонент построена следующая 4-х факторная модель показателей (табл. 1), полностью соответствующая рис.1.

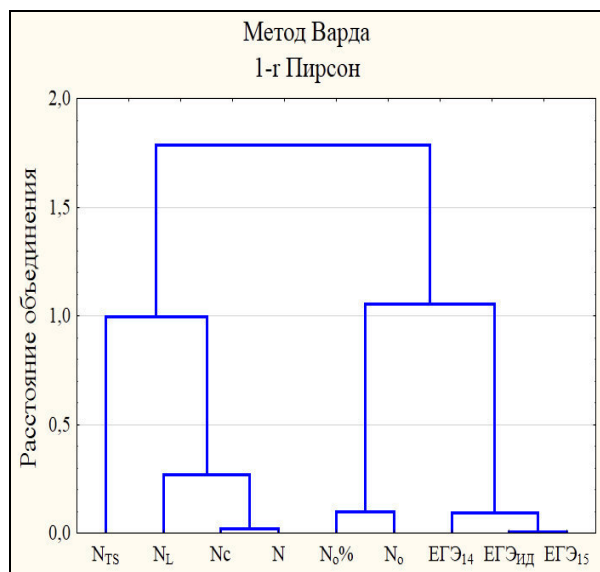


Рис.1. Дендрограмма исходных показателей

Таблица 1
Вращаемые факторные нагрузки в полученной 4-х факторной модели

Перемен.	Фактор 1	Фактор 2	Фактор 3	Фактор 4
ЕГЭ ₁₅	0.350	0.173	0.267	0.092
ЕГЭ ₁₄	0.920	0.222	0.198	0.085
ЕГЭ _{ид}	0.935	0.177	0.258	0.094
N	0.241	0.921	0.193	0.146
No	0.276	0.289	0.902	0.042
No%	0.301	-0.018	0.939	0.053
N _{TS}	0.133	0.167	0.058	0.975
N _L	0.118	0.888	0.126	0.080
N _c	0.201	0.948	0.012	0.097
Общ.дис.	2.892	2.759	1.928	1.017
Доля фактора	0.321	0.307	0.214	0.113

Для дальнейшего анализа была использована 3-факторная модель, включающая тройку первых основных факторных показателей и объясняющая изменчивость на $\approx 84\%$. Из табл.1 видно, что высокие факторные нагрузки первоначальных показателей разделились по факторам таким образом: фактор 1 – фактор вступительных испытаний Фегэ характеризуется положительной корреляционной связью с ЕГЭ₁₅, ЕГЭ₁₄, ЕГЭ_{ид} и интерпретируется как фактор среднего балла ЕГЭ; фактор 2 – фактор Φ_{No} характеризуется положительной корреляционной связью с No и No% и интерпретируется как фактор количества студентов, поступивших с помощью победы в олимпиаде; фактор 3 – фактор численности Φ_N характеризуется положительной корреляционной связью с N, N_c, N_L и интерпретируется как фактор численности поступивших по НПМ.

В построенном 3-х мерном факторном пространстве $\{\Phi_{егэ}, \Phi_{No}, \Phi_N\}$ проведена кластеризация 119-и ВУЗов. В результате получено разбиение 119 ВУЗов на 12 кластеров. Согласно параметрическому F-критерию и ранговому критерию Краскела-Уоллиса кластеры групп ВУЗов различаются высоко значимо по всем факторным показателям (уровень значимости $p < 0,0005$).

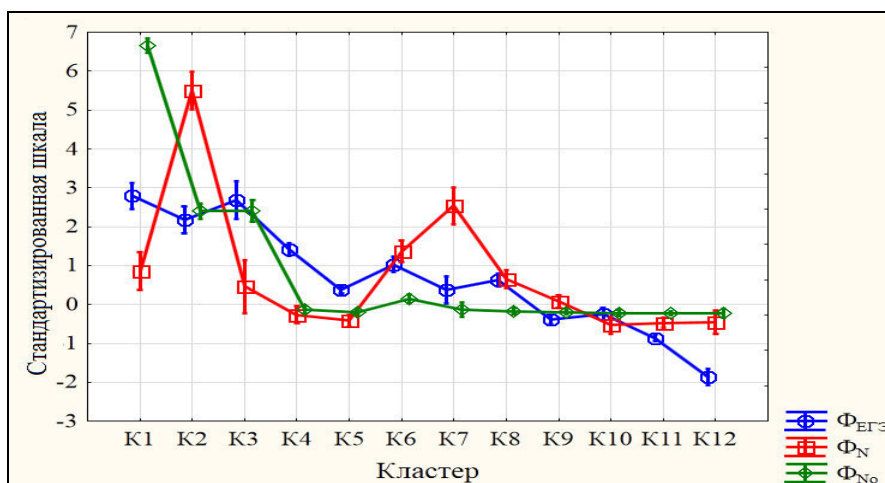


Рис.2. График средних для каждого кластера

Из рис. 2 наглядно видно, что кластеры K4, K5, K7 - K12 являются однородными по фактору Φ_{No} (на уровне значимости $p > 0,1$), причем согласно параметрическому F -критерию K6 отличается от соседнего кластера K7 статистически значимо ($0,005 < p_F \approx 0,015 < 0,05$), ранговый критерий Краскела-Уоллиса сглаживает данное различие до незначимого ($p_{F-Y} > 0,1$). Кластеры K4, K5, K10 - K12 являются однородными и по фактору Φ_N , причем K9 отличается от кластера K12 сильно значимо ($0,0005 < p_F \approx 0,002 < 0,005$) по F -критерию, но статистически значимо ($0,005 < p_{F-Y} \approx 0,016 < 0,05$) по критерию Краскела-Уоллиса. По фактору Φ_{EG} выделяется, например, однородная группа кластеров K1, K3, причем K4 отличается от них по F -критерию сильно значимо ($0,0005 < p_F \approx 0,002 < 0,005$), а по ранговому критерию Краскела-Уоллиса – статистически значимо ($0,005 < p_{F-Y} \approx 0,04 < 0,05$), при этом K2 отличается от K1 статистически значимо ($p_F \approx 0,013$ и $p_{F-Y} \approx 0,046$).

Итоги кластерного анализа ВУЗов по совокупности факторов, учитывая результаты множественных сравнений кластерных средних для каждого фактора, дают возможность провести классификацию ВУЗов в порядковой шкале стандартизированных измерений, считая в качестве уровня «Абсолютный лидер» – стандартизированный интервал ($> +4$), «Лидер» – ($+2; +4$), «Выше среднего» – ($0,75; 2$), «Средний» – ($-0,75; +0,75$), «Ниже среднего» – ($-2; -0,75$).

Используя данную стандартизированную шкалу можно сделать вывод, что, по фактору Φ_{EG} ВУЗы кластеров K1 (ВШЭ, СПб НИ ун-т информационных технологий), K2 (МГУ, СПбГУ), K3 (МФТИ) являются «лидерами», средний балл ЕГЭ данных кластеров превышает 87, а ВУЗы кластеры K12, который включает в себя такие ВУЗы, как Чеченский ГУ, Братский ГУ, демонстрируют уровень ЕГЭ близкий к «ниже среднего», средний балл ЕГЭ которых ниже 50. Абсолютным лидером по количеству подготовки специалистов по НПМ являются ВУЗы кластеров K2 (более 400 студентов), также лидерами являются ВУЗы K7 (Нижегородский ГУ, Воронежский ГУ). Абсолютным лидером по количеству студентов, поступивших по олимпиаде является кластер K1, долевое количество студентов, принятых по олимпиадам данного кластера превышает 70%. Томский ПУ (средний балл ЕГЭ ≈ 73 , $N = 40$) входит в K5, а Томский ГУ (средний балл ЕГЭ ≈ 67 , $N = 150$) – в K8, которые различаются только по Φ_N и характеризуется средним уровнем всех факторных показателей.

Результаты проведенного статистического анализа могут быть учтены при принятии управленческих решений в рамках проходящей реформы высшего образования.

Выводы.

1. На основании корреляционного анализа выявлены 3 группы исходных показателей: $\{EG_{14}, EG_{15}, EG_{ид}\}$, $\{N, N_c, N_L\}$, $\{No, No\%\}$, что позволило с помощью факторного анализа построить 3-х факторную модель, объясняющую изменчивость на $\approx 84\%$.
2. В рамках кластерного анализа в 3-х мерном факторном пространстве $\{\Phi_{EG}, \Phi_{No}, \Phi_N\}$ построена 12-ти кластерная высококачественная модель 119-ти вузов по НПМ. Для каждого факторного показателя выделены группы однородных кластеров и оценен уровень значимости различий неоднородных кластеров.

СПИСОК ЛИТЕРАТУРЫ.

1. Качество бюджетного приема в государственные вузы [Электронный ресурс]. – Режим доступа: <http://www.hse.ru/ege/>. – 25.03.2016.
2. Боровиков В.П. STATISTICA. Искусство анализа данных на компьютере: Для профессионалов. – СПб.: Питер, 2003. – 688 с.